Math 36A: Probability

Denis Patterson, Brandeis

December 19th, 2019

Contents

Chapter 1

Combinatorics

Theorem 1 (Fundamental Principle of Counting) *Suppose we conduct $r \in \mathbb{N}$ experiments and that experiment $i \in \{1, \ldots, r\}$ has n_i possible outcomes. Then there are*

$$n_1 \times n_2 \times \cdots \times n_r$$

possible outcomes of the r experiments.

Corollary 1 *There are $n! := n(n-1)(n-2)\ldots 3.2.1$ ways to arrange $n \in \mathbb{N}$ distinct objects, i.e. when we can distinguish between each of the n objects.*

Corollary 2 *Suppose we have r different types of objects with n_1 objects of type 1, n_2 objects of type 2, and so on, with n_i objects of type $i \in \{1, \ldots, r\}$. If $n_1 + n_2 + \cdots + n_r = n$, then there are*

$$\frac{n!}{n_1!\, n_2! \ldots n_r!}$$

distinct arrangements of the total collection of n objects.

Definition 1 *If $n > k$ with $n, k \in \mathbb{N}$, then we define the binomial coefficients by the formula:*

$$\binom{n}{k} = \frac{n!}{(n-k)!\, k!}.$$

Proposition 1 *Suppose $n > k$ with $n, k \in \mathbb{N}$. There are $\binom{n}{k}$ distinct groups of k objects that can be chosen from a collection of n distinguishable objects, i.e. a distinct group means that the groupings are unique up to permutation of their elements.*

Theorem 2 (Binomial Theorem) *If $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$, then*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Definition 2 *Consider r natural numbers $n_1, n_2, \ldots, n_r$ such that $n_1 + n_2 + \cdots + n_r = n$. The multinomial coefficients are defined by the formula:*

$$\binom{n}{n_1, n_2, \ldots, n_r} := \frac{n!}{n_1!\, n_2! \ldots n_r!}$$

Proposition 2 *If we divide a collection of n distinguishable objects into r groups of sizes $n_1, n_2, \dots, n_r$, i.e. $n_1 + n_2 + \dots + n_r = n$, then there are $\binom{n}{n_1, n_2, \dots, n_r}$ possible distinct groupings.*

Theorem 3 (Multinomial Theorem) *If $x_1, x_2, \dots, x_r \in \mathbb{R}$ and $n \in \mathbb{N}$, then*

$$(x_1 + x_2 + \dots + x_r)^n = \sum_{\substack{n_1, n_2, \dots, n_r \in \mathbb{N}: \\ n_1 + n_2 + \dots + n_r = n}} \binom{n}{n_1, n_2, \dots, n_r} x_1^{n_1} x_2^{n_2} \dots x_r^{n_r}.$$

Proposition 3 ("Stars and bars 1") *Suppose $x_1, x_2, \dots, x_r$ are positive integers and consider the equation*

$$x_1 + x_2 + \dots x_r = n, \quad n \in \mathbb{N}. \tag{1.1}$$

There are $\binom{n-1}{r-1}$ possible solutions to equation (1.1).

Proposition 4 ("Stars and bars 2") *Suppose $x_1, x_2, \dots, x_r$ are nonnegative integers and consider the equation*

$$x_1 + x_2 + \dots x_r = n, \quad n \in \mathbb{N}. \tag{1.2}$$

There are $\binom{n+r-1}{r-1}$ possible solutions to equation (1.2).

Summary Count ways to choose k objects from a population of n distinguishable objects:

	Order matters	Order doesn't matter
With replacement	n^k	Stars & bars
Without replacement	$n(n-1)\dots(n-k+1)$	$\binom{n}{k}$

Chapter 2

Axioms of Probability & Conditional Probability

Definition 3 (Probability Space) *A probability space is a triple* $(\Omega, \mathcal{F}, \mathbb{P})$ *where*

- Ω *is the sample space of the random experiment to be modeled, i.e. the set of all possible outcomes of the experiment,*
- $\mathcal{F}$ *is a collection of subsets of* Ω *(these subsets are called events) to which we may assign probabilities,*
- $\mathbb{P}$ *is a probability measure, i.e.* $\mathbb{P} : \mathcal{F} \mapsto [0,1]$ *or (equivalently)* $\mathbb{P}$ *is a function which maps events to their probabilities.*

Note If Ω is a finite set, then we always simply choose $\mathcal{F} = 2^{\Omega}$ (the set of all subsets of Ω).

Proposition 5 (DeMorgan's Laws) *If* $E_1, \ldots, E_n$ *be a collection of sets, then*

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

where E_i^c *denotes the complement of the set* E_i.

Definition 4 (Axioms of Probability) *Suppose* $\mathbb{P} : \mathcal{F} \mapsto [0,1]$ *where* $\mathcal{F}$ *is a collection of subsets of a set* Ω *which we call the sample space.* $\mathbb{P}$ *is a probability measure if:*

(i.) $\mathbb{P}[\Omega] = 1$,

(ii.) $0 \leq \mathbb{P}[E] \leq 1$ *for each* $E \in \mathcal{F}$,

(iii.)

$$\mathbb{P}\left[\bigcup_{i=1}^{n} E_i\right] = \sum_{i=1}^{n} \mathbb{P}[E_i],$$

for any sequence of sets $\{E_i\}_{i=1}^{n} \in \mathcal{F}$ *such that*

$$E_i \cap E_j = \emptyset \quad \textit{for} \quad i \neq j.$$

Proposition 6 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.*

(i.) For any set $E \in \mathcal{F}$, $\mathbb{P}[E^c] = 1 - \mathbb{P}[E]$.

(ii.) If $D, E \in \mathcal{F}$ and $D \subset E$, then $\mathbb{P}[D] \leq \mathbb{P}[E]$.

(iii.) If $D, E \in \mathcal{F}$, then $\mathbb{P}[D \cup E] = \mathbb{P}[D] + \mathbb{P}[E] - \mathbb{P}[D \cap E]$.

Definition 5 (Uniform Probability Space) *If Ω is a finite set with $|\Omega| = n$, label the elements of Ω as $\{1\}, \{2\}, \ldots, \{n\}$ and choose the probability measure $\mathbb{P}$ such that*

$$\mathbb{P}[\{1\}] = \mathbb{P}[\{2\}] = \cdots = \mathbb{P}[\{n\}].$$

We call this the uniform probability space (on Ω).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with $E, F \in \mathcal{F}$ such that $\mathbb{P}[F] > 0$.

Definition 6 (Conditional Probability) *The probability of E given F is denoted by $\mathbb{P}[E|F]$ and is given by the formula*

$$\mathbb{P}[E|F] = \frac{\mathbb{P}[E \cap F]}{\mathbb{P}[F]}.$$

Proposition 7 *The probability measure given by $\mathbb{P}[\,\cdot\,|F]$ obeys the axioms of probability, i.e. it is a valid probability measure.*

Theorem 4 (Baye's Theorem)

$$\mathbb{P}[E|F] = \frac{\mathbb{P}[F|E]\,\mathbb{P}[E]}{\mathbb{P}[F]}.$$

Definition 7 *A collection of events $\{E_i\}_{i=1}^{\infty}$ is a partition of the sample space Ω if*

$$\bigcup_{i=1}^{\infty} E_i = \Omega \quad \textit{and} \quad E_i \cap E_j = \emptyset \textit{ for } i \neq j.$$

Proposition 8 (Law of Total Probability) *If $\{E_i\}_{i=1}^{\infty}$ be a partition of Ω and suppose F is another event in $\mathcal{F}$, then*

$$\mathbb{P}[F] = \sum_{i=1}^{\infty} \mathbb{P}[F|E_i]\,\mathbb{P}[E_i]$$

Definition 8 *Two events E and F are said to be independent if*

$$\mathbb{P}[E \cap F] = \mathbb{P}[E]\,\mathbb{P}[F].$$

Proposition 9 *If the events E and F are independent, then*

(i.) E and F^c are independent,

(ii.) E^c and F are independent,

(iii.) E^c and F^c are independent.

Chapter 3

Random Variables

Suppose Ω is a finite or countable set and choose $\mathcal{F} = 2^{\Omega}$. Let $\mathbb{P}$ denote a probability measure on $(\Omega, \mathcal{F})$.

Definition 9 (Random Variable) *A random variable is a map from the sample space Ω to some set T, i.e. $X : \Omega \mapsto T$.*

Definition 10 *Given a random variable $X : \Omega \mapsto T$, the law of X is given by*

$$\mathbb{P}[X \in A] = \mathbb{P}\left[\{\omega \in \Omega \,:\, X(\omega) \in A\}\right], \quad A \in T.$$

The family $\{\mathbb{P}[X = j],\, j \in T' := X(\Omega)\}$ is sometimes called the law of X and defines a new probability measure on T' (the range of X).

Definition 11 *As defined above, a random variable X can take on at most countably many values (a finite number or countably infinitely many) and random variables with this property are called discrete.*

The law of a discrete random variable is called its probability mass function (PMF), i.e.

$$f_X(x) = \mathbb{P}[X = x], \quad x \in T'.$$

The cumulative distribution function (CDF) of X is given by

$$F_X(x) = \mathbb{P}[X \leq x], \quad x \in T'.$$

The expectation of X is given by

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega)\mathbb{P}[\omega].$$

Proposition 10 (Law of the Unconscious Statistician) *If $X : \Omega \mapsto T$ is a discrete random variable and $g : \mathbb{R} \mapsto \mathbb{R}$ is any real-valued function, then*

$$\mathbb{E}\left[g(X)\right] = \sum_{x \in T'} g(x)\, f_X(x).$$

In particular,

$$\mathbb{E}\left[X\right] = \sum_{x \in T'} x\, f_X(x).$$

Corollary 3 *If X is a discrete random variable with finite expectation and $a, b \in \mathbb{R}$, then*

$$\mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b.$$

Definition 12 (Variance) *If $X : \Omega \mapsto T$ is a discrete random variable, then its variance, denoted $Var(X)$, is given by*

$$Var(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

3.1 Discrete Random Variables

Notation: We use the symbol "$\sim$" to denote that a random variable has a given (typically named) distribution. For example, $X \sim \text{Binomial}(n, p)$ means that X is a Binomial random variable with parameters n (number of trials) and p (probability of success).

Proposition 11 *If $X_1, X_2, \ldots, X_n$ are Bernoulli random variables with parameter $p \in [0, 1]$, then*

$$\sum_{i=1}^{n} X_i \sim Binomial(n, p).$$

Definition 13 (Bernoulli) *A random variable X is Bernoulli(p) distributed for some $p \in [0, 1]$ if it has a PMF of the form*

$$\mathbb{P}[X = 1] = p, \quad \mathbb{P}[X = 0] = 1 - p.$$

Definition 14 (Binomial) *A random variable X is Binomial(n, p) distributed for some $p \in [0, 1]$ and some $n \in \{1, 2, 3, \ldots\}$ if it has a PMF of the form*

$$\mathbb{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \ldots.$$

Definition 15 (Poisson) *A random variable X is Poisson(λ) distributed for some $\lambda \in (0, \infty)$ if it has a PMF of the form*

$$\mathbb{P}[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

Definition 16 (Hypergeometric) *A random variable X is Hypergeometric(N, m, n) distributed for $N \in \mathbb{N}$, $m \in \{0, 1, \ldots, N\}$ and $n \in \mathbb{N}$ if it has a PMF of the form*

$$\mathbb{P}[X = k] = \frac{\binom{m}{k}\binom{N-m}{m-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \ldots, n,$$

where we use the convention that

$$\binom{n}{k} = 0 \text{ if } k < 0 \text{ or } k > n.$$

Proposition 12 *Suppose $X_1, \ldots, X_n$ are discrete random variables with finite expectation on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

Proposition 13 *Let X be a discrete real-valued random variable with finite variance and suppose $a, b \in \mathbb{R}$. Then*

$$Var[aX + b] = a^2\, Var[X].$$

Definition 17 (CDF) *If X is a real-valued random variable on some probability space, then its cumulative distribution function F_X (CDF) is given by*

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}\left[X \in (-\infty, x]\,\right], \quad x \in \mathbb{R}.$$

3.2 Continuous Random Variables

Definition 18 (Continuous r.v. - informal definition) *We will call a real-valued random variable X a continuous random variable if it can take on uncountably many values and its CDF can be written in the form*

$$F_X(x) = \mathbb{P}\left[X \in (-\infty, x]\,\right] = \int_{-\infty}^{x} f_X(y)\, dy, \quad x \in \mathbb{R},$$

for some nonnegative piecewise continuous function f_X. We will call f_X the probability density function (PDF) of X.

Proposition 14 *Suppose X is a continuous random variable with CDF denoted F_X and PDF denoted f_X. Then*

(i.) $F_X(\infty) = 1$,

(ii.) for any $a \in \mathbb{R}$,

$$\mathbb{P}[X = a] = 0,$$

(iii.) for any real numbers a and b with $a < b$,

$$\mathbb{P}[X \in [a, b]\,] = \mathbb{P}[X \in [a, b)\,] = \mathbb{P}[X \in (a, b]\,] = \mathbb{P}[X \in (a, b)\,] = \int_a^b f_X(y)\, dy,$$

(iv.) for any $a \in \mathbb{R}$,

$$\mathbb{P}[X > a] = \int_a^{\infty} f(x)\, dx.$$

Definition 19 (Expectation/Variance) *Suppose X is a continuous random variable with PDF denoted f_X. The expected value of X is given by*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

Like in the discrete case, the variance of X is given by

$$Var[X] = \mathbb{E}\left[(X-\mu)^2\right],$$

where $\mu = \mathbb{E}[X]$.

Proposition 15 *Suppose X is a continuous random variable with PDF denoted f_X and let g be a function from $\mathbb{R}$ to $\mathbb{R}$. Then*

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx.$$

Definition 20 *A continuous random variable X has a normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ (written $X \sim N(\mu, \sigma^2)$) if it's PDF is given by*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Proposition 16 *If $X \sim N(\mu, \sigma^2)$, then*

(i.) $X + a \sim N(\mu + a, \sigma^2)$, *for any* $a \in \mathbb{R}$,

(ii.) $\alpha X \sim N(\alpha\mu, \alpha^2\sigma^2)$, *for any* $\alpha \in \mathbb{R}$,

(iii.) $\alpha X + a \sim N(\alpha\mu + a, \alpha^2\sigma^2)$, *for any* $\alpha, a \in \mathbb{R}$,

(iv.) $(X - \mu)/\sigma \sim N(0, 1)$.

Proposition 17 *Suppose X is a continuous random variable with PDF f_X and let $g : \mathbb{R} \mapsto \mathbb{R}$ be a strictly monotone (increasing or decreasing) function which is differentiable. Then the PDF of the random variable $Y = g(X)$ is given by*

$$f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|, & y = g(x) \quad \textit{for some } x \in \mathbb{R}, \\ 0, & \textit{else.} \end{cases}$$

Definition 21 (Hazard Rate) *If X is a continuous random variable with PDF f_X and CDF F_X, then the hazard rate function of X is given by*

$$h_X(t) = \lim_{\epsilon \to 0} \frac{\mathbb{P}[X \in [t, t+\epsilon) \,|\, X > t]}{\epsilon} = \lim_{\epsilon \to 0} \frac{\mathbb{P}[X \in [t, t+\epsilon)]}{\epsilon\, \mathbb{P}[X > t]} = \frac{f_X(t)}{1 - F_X(t)}, \quad t \in \mathbb{R}.$$

Remark 1 *Think about X in the hazard rate definition as a random variable modeling the life span of some device. The hazard rate is the instantaneous rate of failure of the device at time t given that it has lived up until time t.*

Proposition 18 *If X is a nonnegative and continuous random variable with hazard rate function given by $h_X(t)$ for $t \geq 0$, then the CDF of X is given by*

$$F_X(t) = 1 - e^{-\int_0^t h_X(s)\, ds}, \quad t \geq 0.$$

Method for Normal Approximations:

If we want to approximate a distribution by a normal distribution, we should match the mean and variance of the normal to the distribution which is to be approximated. For example, if I want to approximate a Binomial(n, p) distribution by a normal, I should take $\mu = np$ and $\sigma^2 = np(1-p)$.

3.3 Jointly Distributed Random Variables

Proposition 19 *Let X and Y be independent discrete random variables with PMFs f_X and f_Y respectively. Then the PMF of $Z = X + Y$ is given by*

$$f_Z(z) = \mathbb{P}[X + Y = z] = \sum_y f_X(z - y) f_Y(y).$$

Proposition 20 *Let X and Y be independent continuous random variables with PDFs f_X and f_Y respectively. Then the PDF of $Z = X + Y$ is given by*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y)\, f_Y(y)\, dy, \quad z \in \mathbb{R}.$$

Proposition 21 *If $X_1, \ldots, X_n$ are independent normally distributed random variables where X_i has mean μ_i and variance σ_i^2, then*

$$X_1 + X_2 + \ldots X_n \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

Definition 22 (Conditional PMF) *If X and Y are discrete jointly distributed random variables with joint PMF $f_{X,Y}$, then the conditional PMF of X given Y is given by*

$$f_{X|Y}(i, j) = \mathbb{P}[\, X = i \,|\, Y = j \,] = \frac{\mathbb{P}[X = i,\, Y = j]}{\mathbb{P}[Y = j]} = \frac{f_{X,Y}(i, j)}{f_Y(j)}, \quad \forall\, (i, j).$$

Definition 23 (Conditional Expectation - discrete case) *If X and Y are discrete random variables with $f_{X|Y}$ the conditional PMF of X given Y, then we can define the conditional expectation of X given $Y = y$ by*

$$\mathbb{E}[X|Y = y] = \sum_x x f_{X|Y}(x, y).$$

Definition 24 (Conditional PDF) *If X and Y are continuous jointly distributed random variables with joint PDF $f_{X,Y}$, then the conditional PDF of X given Y is given by*

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad (x, y) \in \mathbb{R}^2.$$

Proposition 22 *If X and Y are continuous random variables with the conditional PDF of X given Y denoted by $f_{X|Y}$, then*

$$\mathbb{P}[X \in D \,|\, Y = y] = \int_D f_{X|Y}(x|y)\, dx.$$

Definition 25 (Conditional Expectation - continuous case) *If X and Y are continuous jointly distributed random variables with $f_{X|Y}$ the conditional PDF of X given Y, then we can define the conditional expectation of X given $Y = y$ by*

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y)\, dx.$$

Chapter 4

Moment Generating Functions & Limit Theorems

Definition 26 *If X is a (discrete or continuous) r.v., then the n-th moment of X is given by*

$$\mathbb{E}\left[X^n\right], \quad \text{for each } n \in \mathbb{N}.$$

Definition 27 (Moment Generating Function) *If X is a (discrete or continuous) random variable, then the moment generating function of X is given by*

$$M_X(t) = \mathbb{E}[e^{tX}], \quad t \in \mathbb{R}.$$

Proposition 23 *A random variable is uniquely specified by its moment generating function (MGF), i.e. if two random variables have the same MGF, then they have the same distribution.*

Proposition 24 *If X is a (discrete or continuous) random variable with MGF given by $M_X(t)$ for all $t \geq 0$, then*

$$M_X^{(n)}(t)\Big|_{t=0} = \mathbb{E}\left[X^n\right],$$

where $M_X^{(n)}$ denotes the n-th derivative of the MGF.

Proposition 25 *If X and Y are (discrete or continuous) independent random variables, then*

$$M_{X+Y}(t) = M_X(t)\, M_Y(t), \quad t \in \mathbb{R}.$$

Definition 28 *If E is an event, then the indicator function of E is given by*

$$\mathbb{1}_E = \begin{cases} 1, & \omega \in E, \\ 0, & \omega \notin E. \end{cases}$$

Proposition 26 (Markov's Inequality) *If X is a nonnegative random variable with finite expected value, then*

$$\mathbb{P}[X \geq k] \leq \frac{\mathbb{E}[X]}{k}, \quad k > 0.$$

Proposition 27 (Chebyshev's Inequality) *If X is a random variable with finite expected value μ and finite variance σ^2, then*

$$\mathbb{P}\left[|X-\mu|>k\right] \leq \frac{\sigma^2}{k^2}, \quad k>0.$$

Definition 29 *We say the sequence of random variables $(X_n)_{n=1}^{\infty}$ converges in probability to the random variable X if for each $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0.$$

Proposition 28 *If X and Y are independent random variables, then*

$$Var[X+Y] = Var[X] + Var[Y].$$

Theorem 5 (Weak Law of Large Numbers) *Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with finite expected value μ and finite variance. Then $n^{-1}\sum_{i=1}^{n} X_n$ converges in probability to μ as $n \to \infty$.*

Definition 30 (Convergence in Distribution) *Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables with F_n denoting the CDF of X_n for each $n \in \mathbb{N}$. Suppose X is another random variable with CDF given by F_X. We say that X_n tends to X in distribution as $n \to \infty$ if*

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \textit{for each } x \in \mathbb{R}.$$

Proposition 29 *Suppose that $(X_n)_{n=1}^{\infty}$ is a sequence of random variables each with moment generating function M_{X_n} and X is a random variable with moment generating function M_X such that*

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t) \quad \textit{for each } t \in \mathbb{R}.$$

Then X_n tends to X in distribution as $n \to \infty$.

Theorem 6 (Central Limit Theorem) *Let $(X_n)_{n=1}^{\infty}$ be a sequence of independent and identically distributed random variables with (finite) expected value μ and variance $\sigma^2 < \infty$ (finite variance is a key hypothesis). Define*

$$S_n := \frac{X_1 + \ldots X_n - n\mu}{\sigma\sqrt{n}} \quad \textit{for each } n \in \mathbb{N},$$

and let Z denote a standard normal random variable (i.e $Z \sim N(0,1)$). Then

$$\lim_{n\to\infty} F_{S_n}(x) = F_Z(x) \quad \textit{for each } x \in \mathbb{R}.$$

In other words, S_n tends to a standard normal random variable in distribution as $n \to \infty$.